Everyone Can Assist in the Battle Against COVID-19: A CAS Member's Experience

Kwok Ching Ng

The author of this short essay is a member of the State of Vermont COVID-19 modeling team. The main purpose of this essay is to share the author's experience serving in the team and to show how, with some initial investments in learning the basic domain knowledge, any actuary, or person with quantitative and/or critical thinking skills, can add value in the fight against COVID-19 or other pandemics that might arise in the future. The essay also describes the composition of the Vermont modeling team and what it does in assisting the Governor and his cabinet in making COVID-19 policy decisions. Opinions expressed in this essay are strictly the author's.

THE VERMONT COVID-19 MODELING TEAM

As of mid-October 2020, the State of Vermont arguably had the best track record amongst all 50 states since the beginning of the pandemic for the control and mitigation of COVID-19. It had the lowest COVID-19 death rate, the lowest infection rate, and some of the lowest positivity rates amongst all 50 states. "This should be the model for the country, how you've done it," Dr. Anthony Fauci said in a recent video press conference that Vermont Governor Phil Scott held with the state's media. Dr. Fauci also said that being a rural state does not automatically guarantee a better outcome (1) (2). In fact, we could all see how COVID-19 proliferated in the Upper Midwest and other rural states in September and October. The credit for Vermont's success can be attributed to the Governor and his staff, which early on decided to follow science and advice from State Health Commissioner Dr. Mark Levine, State Epidemiologist Dr. Patsy Kelso, and others including the faculty of the University of Vermont Medical School. The Governor held press conferences at least two times a week in which he and the heads of various state government agencies/departments informed the public of what had been happening and the reasons for imposing specific interventions, as well as reasons for subsequent relaxations. Most of all, Vermonters overwhelmingly trust their state government to make the right decisions and largely comply accordingly.

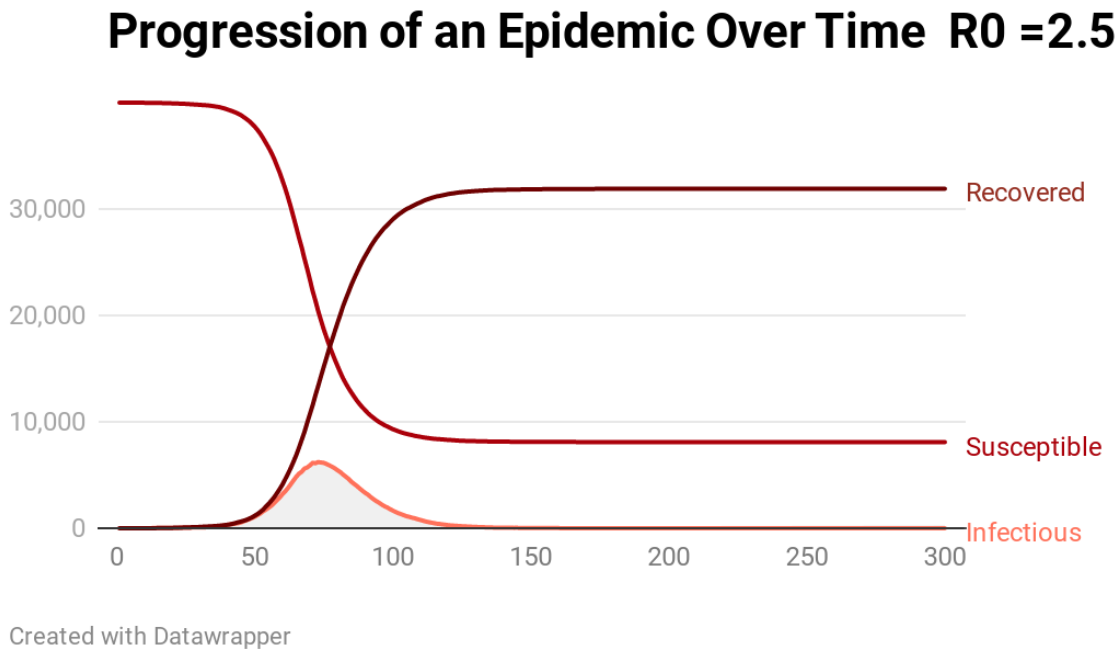
Because all the state's epidemiologists were more than fully occupied with their duties and could not spare time or resources to do COVID-19 modeling, the Commissioner of Financial Regulation, Mike Pieciak, a securities lawyer by training, was asked by the Governor to head up the COVID-19 modeling team in March. A few weeks later, the Deputy Commissioner of Insurance, Kevin Gaffney, who hired me six years before, recommended to Mike that I join his team (while retaining my existing duties). At that time, the only other member of the team was Isaac Dayno, who is a Harvard graduate trained in the liberal arts but with experience in organizing facilities for people with HIV or hepatitis, as well as experience in running political campaigns. As I soon learned, his skillset is important because messaging is a very important part of any public health campaign. A few weeks later, we hired a rising senior from Yale, Ryan Taggard, majoring in Mathematics and Data Science. Ryan's Python programming skills, especially in data collection and graphics, quickly proved to be indispensable in making Mike's weekly presentation to the public come "alive" – with supporting data from around the country. A few weeks after Ryan's arrival, an epidemiologist, Mary-Kate Mohlman, finally joined our team on an as-available basis. Needless to say, having her assistance was a big win. Later still, we began to work with John Adams, the Director of Vermont Center for Geographic Information, to provide us with periodic mobility graphical analyses using SafeGraph mobile phone data. The resolution of SafeGraph's data is an order of magnitude higher than those in Google Mobility Reports.

The roles and tasks of the modeling team evolved over time. From the start, Commissioner Pieciak sought advice from various reputable COVID-19 modeling teams around the country – including Columbia, Northeastern, IHME and Oliver Wyman, among others. All those modeling teams provide state by state projections of COVID-19 confirmed case counts and death counts with intervals of a few weeks to a few months. We would feature one or two modelers' projections in the weekly press conference presentation, along with COVID-19 related Vermont data. Of particular importance were COVID-19 projections under various degrees of non-pharmacological intervention scenarios – such as closure of schools, child-care programs, restaurants, bars, gyms, salons and spas; compliance with Stay Home Stay Safe; and suspension of all in-person business operation, etc., for all businesses and not-for-profit entities.

THE SIR MODEL

Right from the start, I knew I had to acquire as much domain knowledge of epidemiology as quickly as possible. I began to read about SIR models – from basic introductions to dozens and dozens of journal papers in epidemiology. It soon became clear to me that all the COVID-19 non-pharmacological interventions that the country, and in fact the whole world, were considering using can be traced back to, or get their hints from, the SIR models.

I learned that there are many infectious disease models; some are deterministic (SIR, SEIR), and some are stochastic. The most complex ones are network models. The most common one, the SIR model, has been around since the 1930s but still works amazingly well. Common graphical illustration of an epidemic from the beginning to the end based on a simple SIR framework looks like the following:



The model is a dynamic system governed by the following differential equations (5):

$$\begin{cases} \dfrac{dS}{dt} = -\beta \dfrac{S(t) \times I(t)}{N} \\ \dfrac{dI}{dt} = \beta \dfrac{S(t) \times I(t)}{N} - \gamma I(t) \\ \dfrac{dR}{dt} = \gamma I(t) \end{cases}$$

where N is the population size of a closed population, β is the effective contact rate of the disease, γ is the decay rate of the disease, and S(t), I(t), and R(t), respectively, represent the number of susceptible, infectious, and removed (includes those recovered or dead) individuals in the population at time t.

The effective contact rate β is determined by $\beta = \mathrm{T} \times \mu$, where T is the number of exposures per unit of time, and μ is the probability of infection from each occasion of exposure. Clearly, non-pharmaceutical interventions such as social distancing, mask wearing, frequent hand washing, closure of businesses, closure of schools, quarantine, etc., are all efforts to either lower T, or μ, or both, thereby reducing β, which in turn lowers the rate of infection or disease transmission – according to the equation for *dI/dt* above.

I have included some additional information about the SIR model in the Appendix of this essay. There are many journal articles on infectious disease modeling. For example, see (5) for the basic SIR model, (6) for an application of an SEIR model, (7) for an example of a Markovian stochastic SIR model, and (8) for a lengthy discussion of SIR Network models (not an easy reading, though).

A NEW METHOD FOR ESTIMATING ACTIVE (INFECTIOUS) CASE COUNT

At the early stage of the pandemic, measures of an epidemic's reproduction number (commonly referred to as R_0 and R_t) were of great interest to many people. I read papers on the subject and then requested the state's COVID-19 Data Team (who are all epidemiologists) in the Vermont Health Department to provide me with "infector-infectee pairs" data. This information would allow me to create a "Serial Interval Distribution" – one of the input requirements for arriving at estimates of R_t. Basically, the data gives a collection of the number of days between symptom onset of a given infector and symptom onset of the person infected by that infector. That is, the data is a (surrogate) frequency distribution of the number of days required for an infector to pass the virus to an infectee - usually in a family setting or in the workplace. I was provided with 50 pairs (many journal articles use 25 to 40 pairs) and found that a two-parameter Gamma distribution provides a good fit to the data. It should be noted that epidemiologists tend to use Gamma or Lognormal for Serial Distributions. Weeks later I proposed to "off-label" use this distribution as the primary tool for measuring infectiousness by state and counties within a state, turning it into a critical tool for Vermont's travel quarantine policy and for assessing how infectious Vermont's neighboring states are at any point in time.

During the middle of May, when the Northeast states' infection rate began to improve, Vermont Governor Scott wanted to partially open Vermont to tourists from the Northeast. He and his cabinet, with consultation from the Health Department and Commissioner Pieciak, settled on a threshold of 400 "active cases" per million population. Residents of counties with active cases lower than that could travel to Vermont without quarantine. It was then up to the modeling team (us) to figure out how to calculate active case counts. This was not a trivial task. At the time, many states did not publish active cases by county. Even more concerning, we did not know how each state defines/computes active cases, as there was and still is no uniform standard. What we needed was a uniform measuring methodology to be applied to all states and all counties within states. Our own Health Department's procedure is to consider an infected person "active" until 30 days after symptom onset, or sooner if the infected person was clinically determined to have "recovered." In their modeling, Oliver Wyman, on the other hand, simply considers a person "active" during the 14-day period from onset. We were about to adopt

Oliver Wyman's method, but the "Serial Interval Distribution" clearly tells us that an infected person is most infectious during the early days of symptom onset, becoming progressively less so. By the 10th day after onset, the person's infectiousness becomes very low. Counting a person infected yesterday and someone who was infected 10 days ago both as "active," is not a good measuring system for the purpose of differentiating degrees of infectiousness among counties. On the other hand, using the area under the Serial Interval Distribution as "a unit of life-time infectiousness" (or life-time viral load for that person's COVID-19 epoch), a newly infected person would be carrying the highest amount of virus to be shed – 1.00 unit. A person infected 5 days ago would have significantly less "life-time viral load" remaining to be shed, say 0.4 unit. Hence one minus the cumulative distribution (1 - F(t)) of the Serial Interval Distribution (the fitted Gamma distribution) corresponding to the number of days (t) after a person's symptom onset, gives us the life-time viral load unit remaining for that person's COVID-19 epoch. As it turns out, this approach for measuring "active cases" had never been used anywhere before.

Initially, this approach was met with some skepticism, as it should. However, people gradually began to see the merits of the approach. Later, we presented it to two professors at Columbia and Northeastern, respectively, and got concurrences from them. We also got very positive comments from public health modelers in Massachusetts. I also tested the methodology using an SIR simulation model in the public domain (SimInf), and found the two produced quite similar, but not exact, results in terms of number of infectious people remaining at successive points along the timeline. As I read more journal articles, I found that an even more theoretically correct distribution for our purpose would be the convolution of the Serial Interval Distribution and the incubation period distribution. But it has been said, perfect is the enemy of good. One week after I submitted the initial version of this essay to the CAS, we received an unexpected e-mail from Professor Ronald Lasky at Dartmouth College informing us that Dartmouth would like to use our approach for estimating active cases. It was gratifying to see our methodology being adopted by an institution such as Dartmouth. See links in the References section to the documentation of our methodology (3) and our weekly presentation (4).

RANDOM TESTING NOT EFFECTIVE IN IDENTIFYING THE INFECTED BUT ASYMPTOMATIC

Another "discovery" I made during the last several months was, while random COVID-19 surveillance testing could tell us the prevalence of COVID-19 infections in a geographic area, it has a very low expected positivity rate and therefore could be diverting limited resources away from the very important tasks of identifying infected individuals who are symptomatic for quarantine, and from the state to follow up with contact tracing. On the other hand, the positivity rate from testing symptomatic people for COVID-19 is almost always an order of magnitude higher than from performing surveillance testing on people with no symptoms, including asymptomatic individuals. This accidental "discovery" occurred when I was trying to help answer a critic of our active case estimation methodology. Using binomial approximation of Poisson, I concluded that:

With a relatively low population infection rate, conducting random surveillance COVID-19 testing has very little chance of identifying infected individuals, symptomatic or not. In the case of Vermont, the 7,000 tests performed per week in July were expected to only identify approximately two infected individuals or less, per week. Clearly, that would not be an efficient use of COVID-19 testing resources if test kits are in short supply, or future supplies are highly uncertain.

The main reason why Vermont had 30+ to 50+ positive cases identified per week in June to July was due to "self-selecting" or "self-selection" bias either because symptomatic individuals went to get tested/see a doctor, or contact tracing triggered the need for a test.

This highlights the reality that while surveillance testing is useful for determining the prevalence of COVID-19 in a state or a county, it offers very little help in identifying asymptomatic infected individuals for the purpose of stopping them from infecting others. This would still be true even if we increased total testing per week, say fivefold to 35,000, in Vermont. We will have to rely on other means to control spread by such individuals, such as requiring everyone to wear face masks because, by definition, we cannot tell that a person is infected if the

person happens to be of the asymptomatic type. Only testing can tell that. However, we do not have the resources to test everyone who is asymptomatic on a regular basis. The CDC's best guess is that 20% to 50% of all those who are infected are asymptomatic.

I passed on my observations to the team/the Commissioner. A few days later July 24th, the Governor announced a mask wearing mandate - to be effective August 1st. It was something he had been reluctant to do for some time because he preferred educating the public over mandates. Also, during that same week, the Health Department stopped encouraging people with no symptoms to get a COVID-19 test, presumably to conserve test kits. I have no way of knowing the extent, if any, to which my conclusions contributed to those decisions. No doubt, a lot of considerations and inputs from experts and advisers were examined before such key decisions were made, including the fact that at that time COVID-19 was surging along the Eastern Seaboard, marching toward Vermont. The important thing to me was that those were right decisions.

OLIVER WYMAN AS VERMONT COVID-19 CONSULTANT

Even though Vermont experienced only small increases in confirmed COVID-19 cases after the Memorial Day and July 4th holidays, there were serious concerns about schools and colleges reopening as September was approaching. State officials saw elevated risks further down the road from large gatherings during major holidays such as Thanksgiving, Christmas, and New Year's; from visitors during the Vermont ski season; and from winter weather in general, when people stay indoors much more. By the middle of August, Commissioner Pieciak decided to enter into a formal agreement with Oliver Wyman (OW) to provide our team with COVID-19 analyses specific to Vermont in weekly video conference calls. This hugely increased our modeling bandwidth as by that time OW had already been providing COVID-19 consulting services to other entities across the country for several months. Their team consists of a well-known leader in financial industry modeling, an epidemiologist, a medical doctor, consultants from their Health & Life Sciences and Financial Services arms, plus software-programming staff. I found OW's analyses and advice helpful and generally well-thought-out, and innovative at times. I especially like their COVID-19 health risk scorecard by state and their analysis on the necessary conditions and timeline for the country to get back to normalcy. See a link to OW's COVID-19 projections by state (open access) in the References section (9).

In September, like many modelers, OW was developing tools for sizing up the potential impacts from school and college reopenings. OW provided some guidance and relevant information to us, but no projections. As time passed, it was becoming clear to all that, thanks to the efforts of all Vermont school districts, colleges and universities, and support from the state government, Vermont's K-12 and higher education were doing very well in absolute terms and, in comparison to Vermont's nearby states, in terms of having low positive COVID-19 case count per capita. The challenging thing for our team was in securing timely information from all the different school districts, colleges, and universities each week, which we then consolidated and presented at the Governor's press conference every Tuesday. Our focus next turned to Thanksgiving.

MODELING POTENTIAL IMPACT OF THANKSGIVING GATHERINGS

Around mid-November, the Governor's office asked Commissioner Pieciak to review a "COVID-19 Event Risk Assessment Planning Tool" that the Georgia Institute of Technology (Georgia Tech) made available online (10) (11). The Commissioner asked Ryan and me to review the tool and share our opinions with him. In addition, he also wanted to see if the tool could be helpful in analyzing scenarios around Thanksgiving.

My conclusion was that the risk assessment tool relies very heavily on one estimator: The probability that one or more attendees in a large gathering of size N are already infected with COVID-19 when they arrive. That probability is a function of gathering size and the state's COVID-19 prevalence on any given day. However, once the event size reaches 100% for the said probability for a given state, the tool provides no distinctions between all larger event sizes. In particular, it does not provide any framework for estimating how

many new infections could take place during the gathering for progressively larger events, which is a very important consideration for risk assessment and planning purposes.

Conceptually, the missing piece could be approximately modelled by this formula: Expected Number of New Infections During the Gathering = N x P_{SCt} x ARN, where N is the gathering size; P_{SCt} is the probability of an attendee arriving on date t from state S, county C at the event, already infected; and ARN is the "Attack Rate,"- which represents the proportion of people expected to be newly infected during an event of size N. The next step was to estimate P_{SCt}. I chose to estimate that by the number of active (infectious) case count divided by population corresponding to state S, county C, on date t. Active case count came from the methodology described earlier in this essay. Estimating Attack Rates was much harder. Given that Thanksgiving was only seven days away, I decided to rely on empirical data as a starting point and then used an exponential curve in guiding my selections of AR for various sizes N. The empirical data came from various COVID-19 outbreaks traced back to large gatherings such as weddings, church services, birthday parties, etc., reported in the news over the previous few months. The size of the gathering and the number of attendees who got infected were included in those news reports.

For estimating the potential impact of large Thanksgiving gatherings on COVID-19 transmission, Commissioner Pieciak sent me historical surveys by Pew Research Center and YouGov of Thanksgiving family gathering sizes, as well as some less detailed surveys done earlier in 2020. The percentage of such gatherings with guests from out of state was available too. Some of those surveys were done at the national level, and some just for the Northeast Region. Such was the data we could obtain.

After merging and distilling the above information, I estimated the expected new COVID-19 infections due to Thanksgiving gatherings in 2020 - with and without invited guests, based on different assumptions about the extent of voluntary curtailment in inviting non-family members for Thanksgiving. The conclusion: new infections could be 4 to 6 times higher, relative to having no invited guests for Thanksgiving.

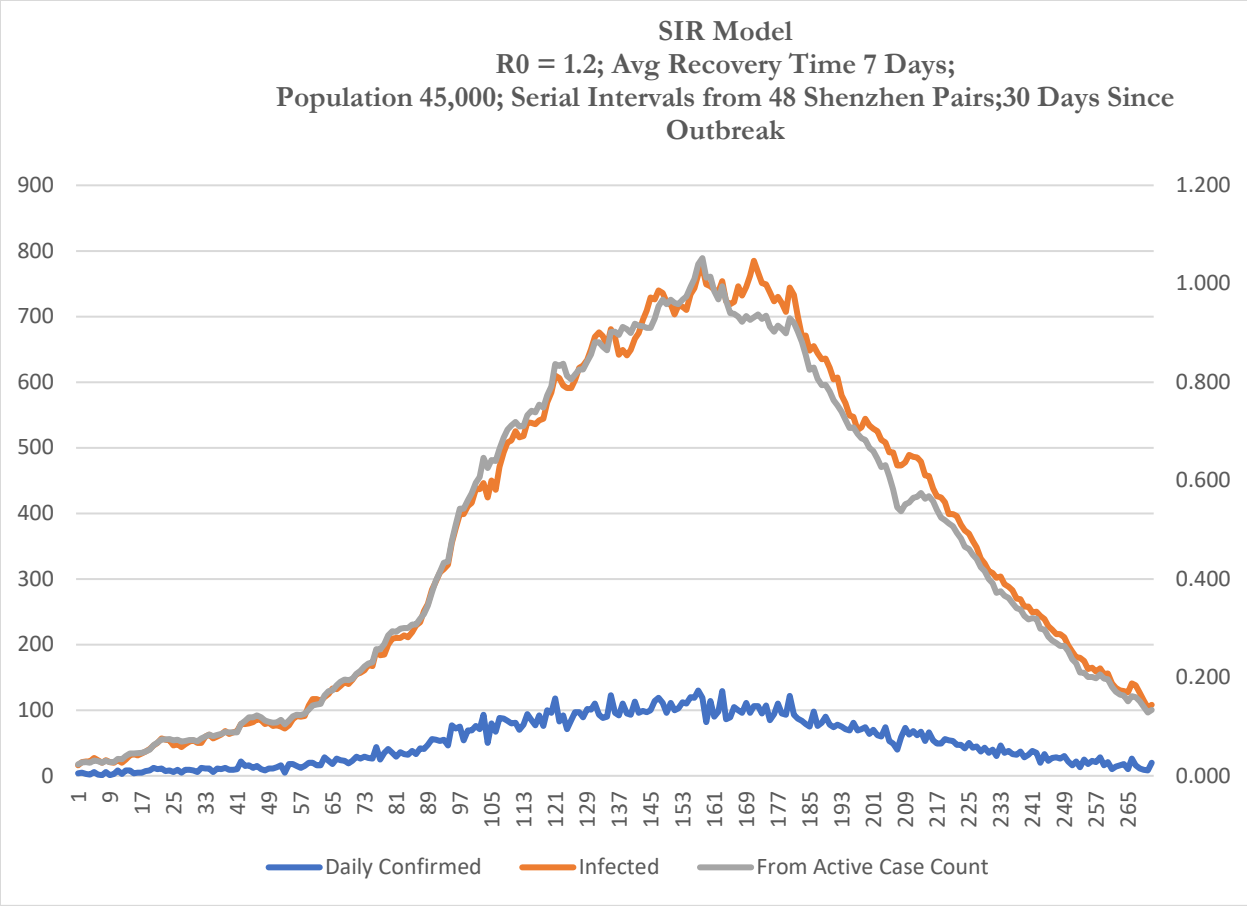
The Commissioner then asked me to send the model and results to Oliver Wyman for feedback. Two days later OW arrived with their own estimates, which turned out to be very similar to ours. Their approach was similar to ours too, except they used tools that they had already compiled/built for estimating the "Attack Rates" more scientifically: room dimensions, ventilation rate, duration of gathering, mask wearing, breathing volume rate, exhalation rate, etc., are in the model. See the link for "COVID-19 Indoor Safety Guideline" by researchers from MIT on how these considerations affect COVID-19 transmissions in the References section (12). Our findings were included in Commissioner Pieciak's presentation at the Governor's November 24th press conference. The intent was to reinforce the message, numerically, that the consequences from large Thanksgiving gatherings could be very severe, with the hope that most Vermont households would keep their gatherings small.

EVERYONE CAN ASSIST IN THE BATTLE AGAINST COVID-19

The modelling I described above could be comfortably performed by most actuaries. An actuary with more advanced skills than me in programming and modeling might have done more. Within our team, neither the Commissioner (a lawyer), nor Isaac, who has a degree in History, have much prior quantitative training. Yet they have a good grasp of the pandemic and are deeply involved with the state's daily battle with COVID-19. Clearly actuaries, statisticians, data scientists, economists, physicists, engineers … all have the capacities to help solve problems on the quantitative side of COVID-19. The world just does not have enough epidemiologists and doctors right now, so do not hesitate to take a deep dive into learning basic epidemiology and then help in any way that you can. We do not know what the future holds.

APPENDIX

Regarding the SIR model, I would suggest that once you have acquired some understanding of how the model works, say from (5), you would want to move on to read (or skim) as many other COVID-19 related articles on epidemiology as your schedule permits. Most epidemiology journals around the world offer free access to everyone during this pandemic. I would also highly recommend trying out the R package SimInf: An R Package for Data-Driven Stochastic Disease Spread Simulations. Vary the input SIR parameters and see how each epidemic unfolds. I used SimInf to test our Active (Infectious) Case Count methodology by running many hypothetical epidemics. The Infectious counts generated by SimInf and our Active (Infectious) Case Counts were in good agreement over the course of most hypothetical epidemics. The following is an example:



REPRODUCTION NUMBER R_0 AND R_t

The most comprehensive paper on Reproduction Number R is in (13) by The Royal Society. You would be better off spending time reading that paper than reading most other papers on R. I quote from the first page of the paper:

"R_0 represents the basic reproduction number, which is the number of secondary infections generated from an initial case at the beginning of an epidemic, in an entirely susceptible population. In contrast, R_t is the reproduction number at time t since the start of the epidemic. As more individuals are infected or immunised, R_t captures the number of secondary infections generated from a population consisting of both naïve/susceptible and exposed/immune individuals and therefore it both changes in value over time and will always be less than R_0."

It is important to know the same virus can have different R_0 depending on where the virus is taking hold. The R_0 for COVID-19 is expected to be different in the metropolitan areas of the US than, say, Fairbanks, Alaska, or Mongolia. This is because R_0 is partly determined by the biology of the virus, and partly driven by how people live and interact.

To estimate R_t, read the paper by Anne Cori et al. (2013) (14). There is also a corresponding online app for estimating R_t called EpiEstim App, and an R package called EpiEstim.

REFERENCES

(1) Dr. Facui spoke in the Vermont Governor's press conference on 9/15/2020: https://www.youtube.com/watch?v=axG1eJfXn00&list=PL-xsDpLCa0iRMj4dQrqum7uoXaQmqeGvT&index=11 (Dr. Fauci spoke 5 minutes into the conference)

(2) Which states had the best pandemic response, https://www.politico.com/news/2020/10/14/best-state-responses-to-pandemic-429376

(3) Vermont Modeling Team Active Case Count methodology: https://dfr.vermont.gov/document/travel-map-methodology

(4) Vermont Modeling Team presentation slides: https://dfr.vermont.gov/about-us/covid-19/modeling

(5) Ulrich Nguemdjo et al. (2020), Simulating the progression of the COVID-19 disease in Cameroon using SIR models: https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0237832

(6) Alberto Gadio et al. (2020), SEIR Modeling of the Italian Epidemic of SARS-CoV-2 Using Computational Swarm Intelligence: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7277829/

(7) Henry Tuckwell & Ruth Williams (2006), Some properties of a simple stochastic epidemic model of SIR type: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7094774/

(8) Leon Danon et al. (2011), Networks and the Epidemiology of Infectious Disease: https://www.hindawi.com/journals/ipid/2011/284909/

(9) Oliver Wyman: https://pandemicnavigator.oliverwyman.com/forecast?mode=states&region=US_US-NY&panel=baseline

(10) Georgia Tech paper on COVID-19 events risk assessment: https://www.nature.com/articles/s41562-020-01000-9

(11) Georgia Tech online app for COVID-19 events risk assessment: https://covid19risk.biosci.gatech.edu/

(12) COVID-19 Indoor Safety Guideline: https://indoor-covid-safety.herokuapp.com/

(13) The Royal Society on Reproduction Number R: https://royalsociety.org/-/media/policy/projects/set-c/set-covid-19-R-estimates.pdf?la=en-GB&hash=FDFFC11968E5D247D8FF641930680BD6

(14) Anne Cori, et al. (2013), A New Framework and Software to Estimate Time-Varying Reproduction Numbers During Epidemics: https://academic.oup.com/aje/article/178/9/1505/89262

ACKNOWLEDGEMENT

The author would like to thank the reviewers of this essay, Sandra Callanan, FCAS, and Denise Ambrogio, FCAS, for their meticulous edits, insightful comments and suggestions.